
Etude de la fonction et des mécanismes d'évolution des séquences répétées centromériques chez les Primates

Sarah Kaddah

Tuteur : Loïc Ponger

Structure et Instabilité des Génomes

MNHN - CNRS UMR 7196 / INSERM U1154 - Sorbonne Universités



Remerciements

Merci à Namrod pour toute la partie sur la bibliographie. Retrouvez ses questions FAQ qui ont permis la rédaction de cette partie.

Merci à f-leb, LittleWhite et Metalman pour leurs conseils et la relecture. Merci à ced et jacques_jean pour la correction orthographique et typographique.

Rapport de stage

Sarah Kaddah

1^{er} mars 2018

Résumé

Votre résumé commence ici... ...

Abstract

Abstract begins here... ...

1 Introduction

1.1 Les séquences centromériques

-> info supp sur l'ADN satellite

Le centromère est une structure chromatinienne caractérisé par la présence de CENP-A. Cette protéine, très conservée au cours de l'évolution, est un variant de l'histone H3. Son rôle est de fixer la position du kinétochore par un mécanisme encore peu connu. En effet, le centromère est le site d'assemblage du kinétochore, un ensemble d'ADN et de protéines. Il joue un rôle important durant la division cellulaire chez les eucaryotes en permettant l'attachement du fuseau mitotique pour la ségrégation des chromosomes. Le centromère et les protéines impliquées sont relativement bien conservés. Au contraire, l'ADN sous-jacent est très diversifié et l'organisation varie d'un taxon à l'autre. Cependant, une caractéristique commune est retrouvée chez toutes les espèces : de l'ADN centromérique répété en tandem nommé ADN satellite. Ces séquences représentent 5% du génome. Les répétitions s'étendent de 7pb à 3,2kb avec des séquences de 145-180kb le plus souvent.

1.2 L'ADN α -satellites

-> première mise en évidence des AS

-> théorie gradient de l'âge

-> travaux sur le gorille à dev

L'ADN satellite chez les Primates est connu sous le nom d' α -satellite. Ces séquences centromériques répétées en tandem, riches en AT, sont issues d'un événement d'amplification. Un monomère a une longueur de 171pb et il peut être répété des milliers de fois. Les monomères peuvent être répartis en famille selon leur similarité, les séquences ayant un taux d'identité supérieur à 70%. Ces séquences ont soit une organisation monomérique soit une organisation en répétition d'ordre supérieur. Dans le premier cas, les séquences d'une même famille sont répétées en tandem. Dans le deuxième cas, une suite de monomères appartenant à différentes familles forme une unité, qui elle est répétée en tandem. Ces séquences peuvent avoir un site de liaison à la protéine centromérique CENP-B reconnaît la CENP-B box, un motif spécifique de 17pb. Cette protéine, qui reconnaît et se fixe sur l'ADN, serait présente chez de nombreuses familles de primates. La protéine pJ α , une protéine peu caractérisée, reconnaît un motif qui remplace la CENP-B box.

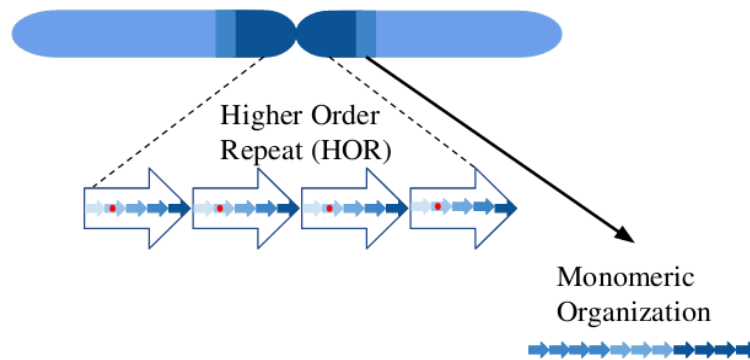


FIGURE 1 – Organisation spatiale des α -satellites. Une couleur correspond à un monomère d'une famille. Les points rouges représentent les sites de fixation à CENP-B.

Les α -satellites ont essentiellement été étudiées chez l'homme. Modèle évolutif avec les centromères en expansion. Une hypothèse concernant l'âge des séquences découle de ces recherches : les séquences les plus récentes apparaissent au coeur du centromères, déplaçant les plus anciennes au péricentromère. D'autres études chez le gorille ont été faites. Le rôle des α -satellites est encore mal connu.

1.3 Le sujet de stage

enchaine sur l'étude chez les cerco, une autre étude de séquençage haut débit

- travaux précédents limités (expliquer pk)
- présentation de mon équipe de leur travail sur les cerco
- la méthode de classification qui a été faite
- Objectif du stage : utilisation de la méthode et comparaison avec des travaux qui ont déjà été faits

Analyser des données de séquençage haut débit en utilisant une méthode de classification identique pour toutes les espèces

2 Matériel et méthode

2.1 Choix des espèces

Les premières espèces choisies sont les *Cercopithecus solatus* et *pogonias*. L'équipe, ayant séquencé leur génome, donnent une description au préalable des α -satellites. Je vais donc appliquer

aux séquences obtenues notre méthode de classification automatisée puis comparer les résultats de ces deux approches.

Le choix des espèces suivantes, le *Macaca fascicularis* et le *Chlorocebus sabaeus*, est fait selon la qualité des reads, obtenus par séquençage haut débit. En effet, les séquences doivent avoir une taille assez longue, soit au minimum 170 pb. Toutes ces espèces sont relativement proches, dans le but de faire des comparaisons inter-espèce par la suite.

2.2 Méthode de classification

2.3 Alignement, phylogénie et consensus

3 Résultat

4 Discussion

5 Conclusion